MORPHOLOGICAL-BASED ADAPTIVE SEGMENTATION AND QUANTIFICATION OF CELL ASSAYS IN HIGH CONTENT SCREENING

*J. Angulo**

Centre de Morphologie Mathématique
Ecole des Mines de Paris
35 rue Saint-Honoré, 77305 Fontainebleau, France

B. Schaack

Laboratoire BioPuces
CEA Grenoble - DSV/ iRTSV
17 rue des Martyrs, 38054 Grenoble, France

ABSTRACT

In fluorescence-labelled cell assays for high content screening applications, image processing software is necessary to have automatic algorithms for segmenting the cells individually and for quantifying their intensities, size/shape parameters, etc. Mathematical morphology is a non-linear image processing technique which is proven to be a very powerful tool in biomedical microscopy image analysis. This paper presents a morphological methodology based on connected filters, watershed transformation and granulometries for segmenting cells of different size, contrast, etc. In particular, the performance of the algorithms is illustrated with cell images from a toxicity assay in three-labels (Hoechst, EGFP, Phalloïdin) on nanodrops cell-on-chip format.

Index Terms— quantitative cytology, mathematical morphology, watershed segmentation, multi-scale gradient, granulometry

1. INTRODUCTION

High content screening (HCS) refers to technological platforms for parallel cells growing in multi-well plates (or in other supports as cell on chip) and fluorescent labelling of proteins of interest (immuno-fluorescence with antibodies, GFP-tagged proteins), together with image capture by automated microscopy and subsequent cell image analysis. HCS is of interest for the discovery of new cellular biology mechanisms (i.e., using siRNA), new pharmaceuticals (i.e., mass screening of potential active molecules) or for the development of new tests for diagnostic/prognostic, for toxicology tests (i.e., evaluation of different compounds at different concentrations).

Cell markers are fluorescent dyes which are extremely specific that will bind to only a few particular proteins in a cell. After image acquisition of a multivariate image by fluorescence microscope (confocal or not), accurate analysis of the morphological changes and quantification of fluorescence parameters require the development of algorithms dedicated to: 1) automatic image processing (segmentation and feature extraction of individual cells and cell populations) and, 2) multi-parametric statistical data analysis and classification. Image segmentation of cells [1] [2] [3] is the most critical step to achieve a robust high throughput system which will be able to segment thousands of cell images without needing a manual interaction. Errors in segmentation process may propagate to the feature extraction and classification.

Mathematical morphology [4] is a non-linear image processing technique which is proven to be a very powerful tool in biomedical microscopy image analysis [5]. This paper presents a flexible morphological methodology based on connected filters [6, 7], watershed transformation [8] and granulometries [4] for adaptively segmenting cells of different size, contrast, etc. We try also to minimise the number of parameters to be set in the algorithms. Three main steps of cell image processing: i) pre-quantification, ii) cell segmentation, iii) quantification, will be detailed in the paper.

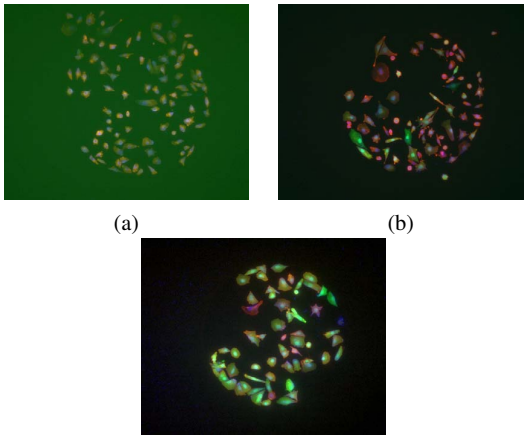

Fig. 1. Three examples of cell images for a toxicity assay: (a) negative control, (b) low concentration of toxic compound, (c) high concentration.

*This work was partially supported by the French Project *GEMBIO-Bioinformatique*, Conseil General des Mines 2004/2006.

In particular, the performance of our algorithms is illustrated with cell images from a toxicity assay in three fluorescent-labels on nanodrops cell-on-chip format [9]. More precisely, it deals with a toxicology screening in HepG2 hepatocellular carcinoma cells: genetically engineered cells to express green fluorescent protein reporter gene GFP under the control of Hsp70 stress promoters [10]. At day 2 the toxic Arsenate is added to the assay and at day 4 the cells are PFA fixated on the chip and Hoechst/Phalloidin labelled before being imaged. Fig. 1 shows three examples of images from this toxicity assay. Phalloïdin is a cytoplasm marker (label against F-actin), f_{cyto}, useful to segment the cell contours (which are however textured and presenting overlapped cells). Hoescht is a nucleus marker (label DNA), f_{nul}, applied for detecting the cell nuclei (markers of individual cells). The intensity of fluorescence of the EGFP, f_{targ}, is used to quantify the effect of the toxics. In Fig. 2 the three fluorescent components of a cell image are given. These images are typical of any HCS experiment, which usually includes a marker for the nucleus, a marker for the cytoplasm and one or several markers associated to the target molecules. It is observed that the cell images present different size and shape, variable contrast and inhomogeneous background. Consequently, automatic image processing algorithms must be able to take into account this variability in order to achieve robust results.

2. PRE-QUANTIFICATION: ESTIMATE OF SIZE/CONTRAST PARAMETERS FOR SEGMENTING

The aim of this preliminary step is to compute a rough estimate of the size/contrast image structures of cells and therefore to automatically fix the parameters of the different morphological filters needed during the subsequent segmentation algorithm. There are two main connected operators for filtering the cell images: the area opening [7], $\gamma^a_{n_a}$, and the contrast opening [6], $\gamma^h_{n_h}$. These operators removes the bright structures which do not verify the criterion associated respectively to the parameter of surface area or of contrast, but without modifying the contours of the remaining structures (i.e., without blurring). In addition the fast implementations using hierarchical queues allow to apply several filters on the same image without time penalty.

The approach of pre-quantification is based on the notion of granulometry [4], i.e., a family of increasing openings or in other words, a series of area (resp. contrast) openings with increasing values of n_a (resp. n_h). Then the pattern spectrum by area is defined as $PS^a(f, n_a) = \frac{\mathcal{M}(\gamma^a_{n_a}(f)) - \mathcal{M}(\gamma^a_{n_a+step_a}(f))}{\mathcal{M}(f)}$, where $\mathcal{M}(f)$ is the volume of the image f; $n_a^{min} \leq n_a \leq n_a^{max}$, n_a^{min}, n_a^{max} and $step_a$ are the size of the minimal area opening, of the maximal area opening and the increasing size of the successive area openings. Typically, we have fixed $n_a^{min} = 100$, $n_a^{max} = 1000$ and $step_a = 100$ pixels. In Fig. 2(d) are given the curves $PS^a(f_{cyto}, n_a)$ for 5 examples of each of the three population typologies represented in Fig. 1 (blue for negative control images, green for low toxic and red for high toxic). Using these curves, and by computing the average of the distribution as well as the slope of the beginning, it is possible to find the "optimal" size of area opening which filters out the texture structures without removing the cells. The pattern spectra by area is obtained for f_{cyto} and for f_{nuc}. In a similar way, a pattern spectrum by contrast is computed for f_{cyto}. In general cases, the pre-quantification can be applied for each image or, working in a training phase, for a selection of cell images from the specific problem and to consider that the estimated parameters of size and contrast operators are then fixed for the rest of images.

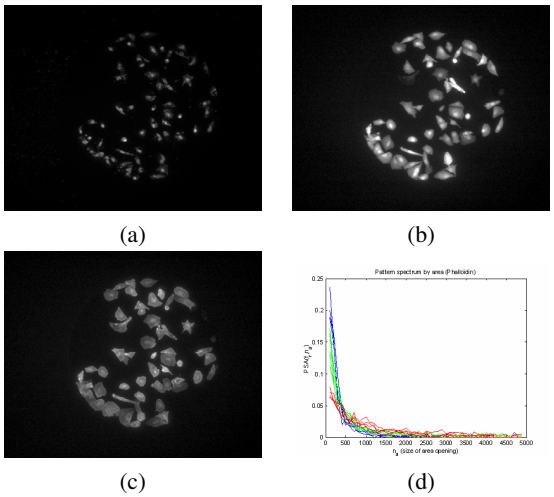

Fig. 2. Splitting the three "colour" components of example (c) from Fig. 1: (a) Hoescht (blue component), f_{nuc}, (b) EGFP (green component), f_{targ-1}, (c) Phalloïdin (red component), f_{cyto}. Examples of pre-quantification: comparative of 15 examples of area pattern spectrum for f_{cyto} (see details in the text).

3. ADAPTIVE SEGMENTATION OF INDIVIDUAL CELLS

The segmentation of individual cells is based on the watershed transformation [8]. Two inputs are needed in order to apply this transformation: 1) an inner marker for each object of interest (here the cells) and a global outer marker for the "background"; 2) a gradient function representing the energy of the contours (here the contours of cytoplasms). In Fig. 3 are illustrated the intermediate images of the different segmentation steps.

Detecting the nuclei (cell markers): The first step consists in removing the non-uniform background by a top-hat

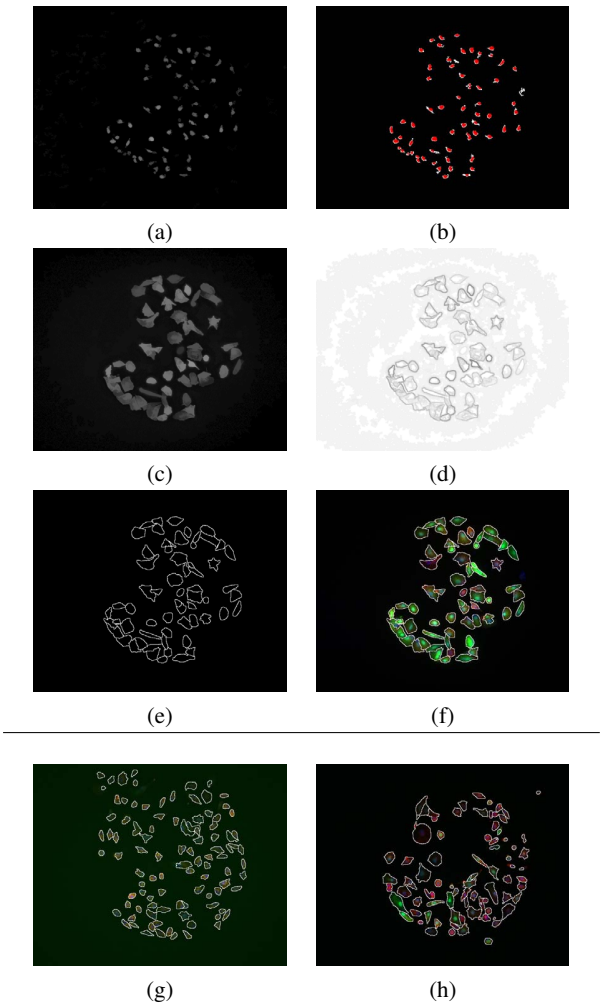

Fig. 3. Top: Algorithm of cell segmentation (see the text for details), (a) f^2_{nuc}, (b) f^3_{nuc} in white and f^4_{nuc} in red, (c) f^2_{cyto}, (d) f^3_{cyto}, (e) f^{cells}, (f) f^{cells} superimposed on original colour image. Bottom: Result of cell segmentation for images of Fig. 1, (g) negative control, (h) low concentration of toxic compound.

transformation: $f^1_{nuc} = f_{nuc} - \gamma_{nbg}(f_{nuc})$, where the opening uses an isotropic structuring element (a disk) of large diameter nbg (typically, 10 times the estimate of cell diameter). The texture of the nuclear structure is removed by an area opening, $f^2_{nuc} = \gamma^a_{ntext}(f^1_{nuc})$, where $ntext$ is smaller than the estimate of nuclear surface area (typically, 75%). The image is then thresholded to obtain a binary mask of nuclei, $f^3_{nuc} = Th_{[u,max]}(f^2_{nuc})$ (the step is not critical on this filtered image and the value is fixed for instance to $u = 5$). The nuclear markers are finally achieved by computing the maxima of the distance function [4], $f^4_{nuc} = Max(\gamma^h_{nsize2}(dist(\gamma^a_{nsize1}(f^3_{nuc}))))$. Note that, before obtaining the maxima, the binary image is previously regularised with a small opening and the distance function is also filtered with a contrast opening. The parameters of both transformations are not critical and can be easily determined (i.e., $nsize1$ is the minimal diameter accepted for a nucleus and $nsize2$ is the minimal radius for a nucleus when overlapped to another bigger one).

Segmenting the cytoplams (cell contours): The preprocessing of the image f_{cyto} is quite similar to that for the previous one. The fluorescence uneven background is removed by a standard top-hat: $f^1_{cyto} = f_{cyto} - \gamma_{nbg}(f_{cyto})$. The texture and irregularities of cell structures are simplified by a sequence of an area opening followed by a contrast opening: $f^2_{cyto} = \gamma^h_{nstrh}(\gamma^a_{nstra}(f^1_{cyto})$ (both size and contrast parameters, $nstra$ and $nstrh$ are estimated respectively from $PS^a(f_{cyto})$ and $PS^h(f_{cyto})$). In mathematical morphology, a gradient is obtained as the different between an erosion and a dilation with an isotropic structuring element of size 1, i.e., $\varrho(f) = \delta_1(f) - \varepsilon_1(f)$. The direct computation of the gradient on the image f^2_{cyto} is not useful for watershed segmentation because, even after the filtering steps, the variations of fluorescence inside a cell are often greater than the variation associated to the contours between cells. In order to solve this problem, we propose to compute a multi-scale gradient, which is obtained as the sum of the gradient for successive contrast-based filtered versions: $f^3_{cyto} = (\varrho(f^2_{cyto}) + \varrho(\gamma^h_{nstrh}(f^2_{cyto})) + \varrho(\gamma^h_{2\times nstrh}(f^2_{cyto})))/3$. This gradient is more regular than $\varrho(f^2_{cyto})$ and contours between cells are enhanced. Note that we can not calculate only the gradient $\varrho(\gamma^h_{2\times nstrh}(f^2_{cyto}))$ since low contrasted cells do not appear after this strong filter.

This gradient is then used for the watershed, $grad = f^3_{cyto}$. The image f^4_{nuc} corresponds to the inner markers for the watershed transformation, mrk_{in}. The background marker will be obtained from the complement of a dilated version of the thresholded cells: $mrk_{out} = [\delta_5(Th_{[3,max]}(f^2_{nuc}))]^c$. The final marker image is the binary image supremum of both images: $mrk = mrk_{in} \vee mrk_{out}$. The final contours of the cells are finally obtained by watershed segmentation: $f^{cells} = wshed(grad, mrk)$. In Fig 3 are given the results of cell segmentation for the current examples. The same algorithm have used for segmentation hundreds of cell images.

We have evaluated precisely the algorithm on 15 representative images, by manualy counting the number of correct segmented, overcut or undetected cells and the average error (i.e., proportion of wrong segmented cells in one image) is between 5% and 8%. The fluorescence images associated to the protein target, like f_{targ} in our examples, can be also used to help the segmentation of the cytoplasm. In fact, in many HCS applications, there is not specific cytoplasm marker and it is indispensable to use the target images for cell segmentation. In our particular problem of toxicology, we have evaluated also the interest of combining a gradient f^3_{cyto} and a similar algorithm on f_{targ} and in most of cases the segmentation is improved.

4. QUANTIFICATION: FROM PARAMETERS OF INDIVIDUAL CELLS TO POPULATION PARAMETERS

Once the individual cells have been segmented, all the parameters of fluorescence intensity from the target images can be computed for each cell. In addition, all the standard parameters for describing the cell morphology: size/shape of cytoplasm, chromatin texture of the nucleus, etc. can be also calculated. We can also compute more global parameters, associated for instance to the cell population. In particular, using morphological granulometries by isotropic openings/closings, it is possible to calculate from the binary image of segmented cells, f^{cells}, the distribution of size/shape of cells (according to the openings) and the evolution of the spatial aggregation of cells (according to the closings). In Fig. 4 are given the results for a selection of 5 images for each studied phenotype. As we can observe, using the positive part of the pattern spectrum (openings), we can easily identify the three phenoytpes. The negative part of the curves (closing) is not meaningful in this problem since the culture is done in 3D nanodrops and there is no phenomenon of aggregation.

5. CONCLUSION AND PERSPECTIVES

We have presented an algorithm for accurate segmentation of fluorescence-labelled cells in HCS applications. A step of pre-quantification allows to automatically setting some parameters of the algorithms in order to eliminate the need of any human interaction. We have also given an example of morphological descriptor associated to the cell population which is useful for phenotypic screening. The developments have been illustrated with results from a toxicology assay using nanodrops cell-on-chip format. We are presently using this methodology for analysing the cell images of other HCS projects, using different types of cells, different target markers, etc. We are also enriching the segmentation methodology by associating to the unsupervised watershed algorithm prior information about the models of cell shape.

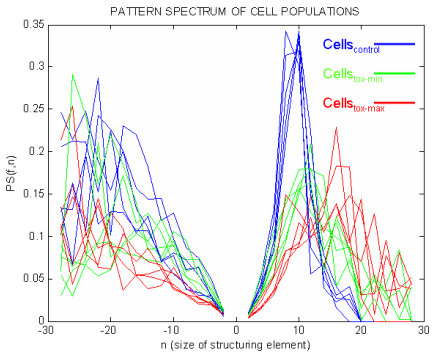

Fig. 4. Examples of population description using granulometries (i.e., pattern spectra by opening/closing) of five examples of segmented cells for each phenotype.

6. REFERENCES

[1] P.M. Kasson, J.B. Huppa, M.M. Davis, and A.T. Brunger, "A hybrid machine-learning approach for segmentation of protein localization data," *Bioinformatics*, vol. 21, pp. 3778–3786, 2005.

[2] B. Neumann, M. Held, and U. Liebel et al., "High-througthput rnai screening by time-lapse imaging of live human cells," *Nature Methods*, vol. 3, pp. 385–390, 2006.

[3] A.E. Carpenter and T.R. Jones M.R. Lamprecht et al., "Cellprofiler: image analysis software for identifying and quantifying cell phenotypes," *Genome Biology*, vol. 7, pp. R100, 2006.

[4] J. Serra, *Image Analysis and Mathematical Morphology. Vol. 1.*, Academic Press. London, 1982.

[5] J. Angulo, J. Klossa, and G. Flandrin, "Ontology-based lymphocyte population description using mathematical morphology on colour blood images," *Cellular and Molecular Biology*, vol. 52, pp. 2–15, 2006.

[6] M. Grimaud, "A new measure of contrast: the dynamics," in *Image Algebra and Morphological Image Processing III*. SPIE, 1992, vol. 1769, pp. 292–305.

[7] L. Vincent, "Grayscale area openings and closings, their efficient implementation and applications," .

[8] S. Beucher and F. Meyer, "The morphological approach to segmentation: The watershed transformation," in *(E. Dougherty Ed.), Mathematical Morphology in Image Processing*. Marcel Dekker, 1992.

[9] B. Schaack, J. Reboud, and S. Combe et al., "A 'drop-chip' cell array for high throughput dna and sirna transfection combined with drug screening," *NanoBiotechnology*, vol. 1, pp. 166–194, 2005.

[10] F. Lemaire, C.A. Mandon, and J. Reboud et al., "Toxicity assays in nanodrops combining bioassay and morphometric endpoints," *PLOS ONE*, vol. 1, pp. e163, 2007.